Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-69

Okay, with that kind of see background right. Now assume that the essential matrix is known, see I mean the fundamental matrix right we know essential matrix is known. How do you know this? Because of the fact that right you can actually compute the say fundamental matrix I mean you can do the 8 point algorithm. 8 point does not mean you have to restrict to 8 points right like I said due to noise you can take more number of points and then and then you can solve for f. So 8 point algorithm gives f and then what you do is you know you know that right f can be found only up to a scale right which is why the norm of f is 1. But then you know last time right if you can say recollect to get f we imposed a rank 2 constraint on it right.

But then in order to find the essential matrix see there is actually a problem, what is the essential matrix by the way Es Tx * R T X * R right this is the essential matrix if you can see if you sort of go back and you know recollect. So in a sense that what was f, f was like Kr – T and then you had T right X R and KL inverse right this is what you had as your f right. So your E is this guy, so E is really Kr what is it Kr transpose fKL right and that gives you E right Tx T X * R not x T X * R right and this is your E. And what this means is that but then when you actually take this f and you assume that assume that essential matrix is known actually means that the camera calibration matrices are known that means the intrinsic is known to you okay.

So you use your 8 point algorithm take the correspondence is solve for f and once you solve for f right you can pre multiply and post multiply but then when you solve for f right you do not really reduce it to rank 2 right at that time I mean right from this see if you want the f itself then you will normally use the rank 2 constraint. But normally when you want to arrive at the essential matrix you let the f as it is right as you kind of get it from get from there because the essential matrix actually has to satisfy something right this is actually a product of a skew symmetric matrix which is skew symmetric in this? T X. Yeah T X is skew symmetric and actually a rotation matrix okay and a rotation matrix and you see by the way every orthogonal matrix is not really a rotation matrix by the way okay see for example if you have some matrix which satisfies a transpose is equal to a transpose a is equal to identity right it does not mean that this automatically you know is automatically a rotation matrix. Rotation is something more specific so for example if I take right determinant of a transpose that is what determinant of a * a no actually it is a little tricky determinant of a transpose but determinant of a transpose and a are the same therefore this becomes determinant of a the square and this should be equal to determinant of identity right that is 1 therefore it means that a determinant of a can be actually + - 1 okay it can be + - 1 but then there is something called a proper rotation versus an improper rotation okay for a proper rotation this should be 1 for a proper rotation determinant of a should be equal to 1 therefore it is not true that any orthogonal matrix becomes a rotation that is not true okay what do you mean by I mean I do not want to go * the detail but improper rotation actually means a reflection followed by rotation but in our terms of rotation we mean and we take an axis about which you rotate there is no reflection and all okay. So proper rotation means that right determinant of a is equal to 1 okay and therefore in all of this right this r that we are putting there is actually a rotation matrix which means that if you compute you know dEt of a of r that will be 1 okay so what this means is right so skew symmetric means that Tx or T X is equal to - whatever right T X transpose okay that is a transpose and T is a vector and rotation means rr no sorry rather determinant of r okay rr transpose equal to identity and determinant of r is equal to 1 so it is a proper rotation matrix.

Now what you do is you take this f right and as computed through the 8 point algorithm then you multiply it by kr transpose and kl right and you can show that skew symmetric you know 3 X 3 skew symmetric matrix will actually have okay will have skew symmetric matrix has 2 identical singular values and you know that it has it is rank 2 identical singular values and a third singular value which is 0 okay this you can show okay but the Tx form you have seen know how it looks I mean you can I mean write any Tx for that matter right if you take with this skew symmetric it will have singular values that are same I mean 2 singular values that are same identical and then third one is 0 as 2 identical singular values and a third singular value which is 0 and the third singular value which is 0 okay and multiplication by a rotation matrix does not change the you know singular values okay therefore E has the same property. So E should ideally have 2 singular values that are identical unlike the f right f we did not have this constraint but then E is coming from this Tx * r and therefore E has an additional constraint on it right which is that the singular value should be identical okay and therefore what is done is if you do what you call right if you do a decomposition so you know this guy right so this is like your E right but then this is not exactly E because if you do this right this may still not give you identical singular values. So if you get a decompose this as U diagonal whatever right V transpose okay if you do this let us call this E dash okay this is still not E okay I will not write this as E I will still call this as E dash so where this diagonal it typically will have some λ 1 0 0 0 λ 2 0 0 0 λ 3 because we have not reduced f to rank 2 okay and then what you do is it can be shown that the best matrix that actually you know minimizes the Frobenius norm between E and E dash there is actually a theoretical proof for this the best E that approximate E dash is the one that has that you can find this let us say U right this is diagonal dash V transpose where diagonal dash is sigma 0 0 0 sigma 0 0 0 0 and sigma what do you think it is λ 1 + λ 2 by 2 okay. This I am not showing the proof but there is a proof that the best E right that you can the best E that approximates E dash in the Frobenius sense is this guy right which with 2 identical singular values derived as λ 1 + average of the 2 Eigen values of coming from E dash okay which one λ 1 λ 2 yeah they are all ordered yeah they are ordered yeah so it is like λ 1 is bigger than or equal to λ 2 is bigger than or equal to λ 3 this is all ordered okay so λ 1 greater than or equal to λ 2 greater than or equal to λ 3 okay. Then now the point is okay now we want to know right how do we actually estimate the pose so the point is this right what we have with us what do we have with us so we have E but then E is actually a decomposition of T x and R right it is actually a product of T x and R but then we have only we seem to have only E with us now right we do not have we do not have access to T and R directly correct so we are kind of looking at a situation where the motion of the camera is unknown and therefore it is going to bring in some ambiguities in terms of what you can estimate.

See for example if I told you the simplest case that there is no R but there is a translation okay there is a translation but then if I do not tell you what is the baseline what will happen if I do not tell you the baseline right I tell you that there is a pure translation there is no this one rotation see till now we had assumed a simpler case where I said that the baseline is given therefore Z is simply right F * baseline upon what F * B by this one a disparity right that is what stereo is right that is how you computed Z last class we showed that but now right we are actually involving it we are kind of making it a making it a little more involved now and we are now saying that suppose I did not know the actual translation that means the baseline I do not know but I know that it is a pure translation now that means that there is there should be some uncertainty now that should come in in terms of what you can estimate can you now estimate Z by an absolute factor this is called reconstruction to an up to an ambiguity okay what is the ambiguity right if you actually if you think about it see it is like this right what information do you have you have only a disparity you do not have any other information right okay intrinsic are known that is all fine you have used that intrinsic you have applied it also you got your E okay now if I look at two images you have a left right pair and I know that a pixel has moved from here to there I have a feature correspondence also right I have done the search I know that right here is the correspondence now this disparity if it turns out to be 5 pixels let us say right if it is a disparity 5 pixels I want to convert it * Z now the problem is I could get this 5 in two ways one could be that the scene is very far away that means this point this depth right which is triangle which I am kind of triangulating that could be very far away but my camera moved a lot right because my camera moved a lot I got 5 pixels or it could be that the point is actually very close to my camera and my and I hardly moved I just moved a little bit and I still get 5 pixels right both can happen no right because of this effects by Z that is what is at play right so when there is a Z and when there is a baseline that is not known so the for the so the so because you do not have access to the camera motion nobody is giving you see there are these papers that talk about how to use the use the inertial sensor that is built * a camera to kind of tap * the motion and all right that is like that is you know a different thing where you can actually use some information it could be noisy but you can plug it in and do something but let us say we do not have anything else right in this case we are assuming that I just have two images with me disparity you can compute but after disparity I am not sure because it could happen that the same disparity I can get for a point that is very far away and large camera motion or for a point very close to the camera and a small camera motion either case I get the same disparity. Therefore when you compute Z it can only be computed up to a scale ambiguity because this translation is something that always translates * something like a scale ambiguity right and there are higher levels of ambiguity as you can already think because here I have assumed rotation is not is not there right but what is there is a rotation then there is going to be additional ambiguities and so on. Therefore what you should remember is when you take structure from motion when you are computing structure from motion using a camera even if it is calibrated even in an uncalibrated camera you can imagine it is going to be even more difficult this is all a calibrated camera right I am using F from F I am going to E and therefore I am reducing my uncertainty to only say T and R now correct and if there was K also was unknown then it means that it becomes an uncalibrated camera case which is even more ambiguous but all of this right that is why when you see a depth map somebody right presents you a depth map do not think that the values of Z that you are getting there are actually is actually a true distance it is not like from there to there it is a it is a it is right you know it is so many millimeters or so many centimeter need not be but a structure may look what you might want it to look like but again right there are still certain certain see right tricky things here which I will which I will talk about as we go along but this wanted to alert you to the fact that we are no longer in the in the same sort of rate a regime where where we had a simple situation baseline given Z in absolute terms all that is going away now right we are going * more complex case now. Now the point is right now we want to now our goal is to be able to is to be able to get get is a T and R right only if you why why do you why do I need is a T and R because finally if I want to triangle it I need my P a projection matrix a camera projection matrix you know is K * R slash T right that is a camera projection matrix and only if I have the camera projection matrix can I do any kind of triangulation so I need my R and T right nobody nobody has given me that yet I have got only the fundamental matrix from there I have come to E now I need T and R so that I can go to my P okay. So so so we need to get a decompose E now okay now define okay now write now I am going to write ES U look at the same U right that you had there and then 1 0 0 0 1 0 0 0 0 V transpose and this I can write because E anyway can be found only up to a scale factor so that sigma I just scaled it off right because this is only up to a scale factor so it is okay to I mean there is no point unnecessarily dragging this sigma along okay.

Now define define Z equal to 0 1 0 okay this is the something red which you will which you which I will tell you why we need this 0 0 what kind of a matrix is this is actually skew symmetric by the way - Z transpose and define another matrix W which is like 0 - 1 0 1 0 0 0 0 1 okay and you can and you can verify that W W transpose is identity and determinant of you see W is equal to 1 that means this is actually this is actually a rotation matrix okay. Now clearly right clearly right what you can show is if you multiply Z * W right I am not going to show this very simple right you can just multiply the 2 matrices you can show that this 1 0 0 0 1 0 0 0 0 and if I do Z * W transpose we need this for a for a for a result that comes later is equal to - of 1 0 0 0 1 0 0 0 0 okay and that why we have this form is because we know that our E is like product of a skew symmetric and actually a rotation matrix right that is the way that is the reason why you are also looking at Z and W of a similar form okay. Now if we decompose right if we decompose ES ES E is equal to some S * R where I mean S is some skew symmetric matrix and R is R is a rotation matrix then one can show that there are there are 2 kind of C choices 2 possibilities okay what are those the possibilities are as follows E is equal to S1 R1 or S2 R2 okay by which what I mean is S1 where S1 is equal to - U Z and this U Z and all is all kind of coming from there only okay from E okay this is not coming arbitrarily Z and W are right as we said what they are U Z W and R1 is equal to what is this U W transpose V transpose S2 is equal to U Z U transpose my U and V are looking somewhat similar but here it is all U Z U transpose okay and R2 is equal to UW V transpose okay now you can actually clearly show right see if you do S1 R1 what will you get will get - U Z U transpose then you get UW transpose V transpose right and then U transpose U is identity you get - U Z W transpose V transpose but I told you that Z W transpose has this form - of that 1 0 0 0 1 0 right. So once you plug that in then you will get the form of V which is like U 1 0 0 and then what is it V transpose right and similarly you can show that S2 R2 will also get a workout to be workout to have the form of AE right not A of E which is like U then that matrix followed by V transpose what is that matrix 1 0 0 0 1 0 0 0 0 okay so you have kind of 2 options now okay now and of course and it is easy to show that right S1 is equal to - S1 transpose so S1 is again skew symmetric S2 is equal to - S2 transpose is also skew symmetric I mean all this you can show just take the transpose and show that all of those holds and then the determinant of R1 is equal to 1 R1 transpose is identity all that follows okay now that and all is not really is not really is irrelevant. Now what we want is now we want to look at how do I get to my translation now to find the translation right what you can do is the following see this S is supposed to be my translational vector right here S * R S is supposed to be skew symmetric kind of translational matrix which is the X of T the even translation right so to find the translation vector okay what can I say I can say that S right S dot T okay which is the same as T X T right this should be equal to the I mean 0 vector right because S is after all S is supposed to be my translation and my translation will say T therefore S should ideally be equal to T X and T X T should be 0 right that is T should be in the say null space of should be in the null space of say S right whatever is that S right I mean so my T should be in the null space of S and so is actually actually λ times T right I mean if T is in the say right null space of S then so is λ T right this we are saying because of the fact that translation can be found only up to a scale factor because λ right it does not mean that this T is now absolute so this T if T works then you see λ T is also true right therefore you do not get a sense for T absolute T there is always be the scale ambiguity so the λ we are throwing it in just to highlight the fact that there is a scale ambiguity okay in translation okay now S dot T is 0 you are not able to why because this S is after all I mean S is what S is supposed to be supposed to be TX right in E is what TX * R correct now S is supposed to be X right T X right is S and therefore it whatever S you find it is I mean if you take S * C T I mean if you take S to be T X then T X T is like taking this one right I mean a X product of the same this one vector with itself no.

No whatever T you find see for example I mean whatever S you get right if you find this in null of S that will give you some vector right we are saying that the vector is not the absolute T it can be get λ times T that is what we are saying it is not the absolute T right whatever we are seeing it is only up to a scale factor that is what we are saying we are not saying that we are getting T so that is why I am saying that it is not λ I right this is λ T okay so if T is valid so is λ T is that still a problem okay now see right TX dot T okay now so this is a sign see the magnitude of λ is not a problem because as I said it not a problem in the sense that we cannot find it find it right you know in any case because it is an unknown right like I said you know because of the Z sort of you see sort of you know camera motion ambiguity but this sign is important okay this scale we cannot do much about it but then we can use a sign to do something okay the sign of λ is important okay while the scale is not in our reach the sign of λ is actually important okay so whether it is actually negative or not or whether it is negative or it is actually positive quantity okay this I will tell you right why this becomes kind of say relevant okay. So what this means is okay now since that S dot T right is equal to 0 okay one can show that right what is that T what is that T right which is going to be in this in null of S so we had let us say what is our S right I mean S was what did we write it as U Z U transpose right what is the form of S I think it was some U Z + - there I know but I am just U Z U transpose right. So if you take right and those ambiguities will come now okay because of the fact that it can be right - I mean S1 S2 and all was there no we will come to that but the first thing is S is equal to U Z U transpose and you can show that and you can show that right the T is actually U3 that means if you do if you actually multiply S with U3 right that is like U Z U transpose U3 vector and U3 is actually the third column of U third column of U okay therefore when you take U transpose right you will get U Z and because it is a third column right so it will be like 0 0 1 right U3 transpose with itself will be 1 no because U transpose U is identity right and if you use the fact that Z has that form right Z had a particular form right whatever what I mean I had thing I had already written 0 1 0 - 1 0 0 if you put that right here you will get U 0 0 0 if you use the U Z the form of Z as I had actually written there and then right this will be equal to 0 if you put U2 and U1 right it would not be 0 okay I mean you would not get this vector as a 0 if you use any other if you use U1 or U2 right you would not find that you know find this vector going to 0 and therefore red T is actually U3 right in a sense okay but then the thing is right we still not sure about the sign and so on okay like I said at the scale the sign right we are not sure. Now so what actually happens is we get 4 solutions now because of the sign ambiguity additional 4 possible solutions for the camera matrix or a projection matrix R T this is what we want right eventually this is what we want R and T because then we can triangulate so what are those so we can have PR right if it is a right camera right which we are talking about let us assume that the first camera is where the coordinate system is right then PR right is let us say these are your choices okay UW V transpose U3 or UW V transpose – U3 like I said at scale we cannot help but then the sign can be - or positive or see I mean you had here no yeah see I mean from that S1 R1 and they say S2 R2 right that you had so from that you have the other one is UW transpose V transpose right because here these are all options of R, R does not have a - right and then again let us say U3 or you have UW transpose V transpose but then – U3 this is V okay V transpose – U3 is this clear correct because of the sign ambiguity in U3 I mean you can find out U3 but then it is up to a sign ambiguity okay and of course you know there is also scale ambiguity okay so right it could be actually it could be actually right any of these 4 now the point is right which one to pick now here is where there is a principle called this one chirality okay I mean I do not know the exact origin of this word one should check I think so what this means is chirality what this means is means is right do you know that that particular scene point is in front of the camera or to the back of the camera see when you solve for this no what will happen you will see suppose you use all 4 right suppose I use the first option and then I do a triangulation and I still I am not shown you how to actually do the triangulation but then assume that you can do the triangulation right and you kind of find your find your X tilde the capital X tilde right which is what you are interested see what I mean is if you go back to your equation which is like X tilde dash is equal to what is this PR PR * X tilde right now I am saying that if you use your PR and somehow if you are able to triangulate we have not shown how to do that from here of course we cannot do but there is a way to actually do it so if you can do triangulation then you will get an X tilde right though the Z sitting at X tilde can be a positive or actually it can be it can have this one you know a negative value that the depth right now if it is a negative value it means that the scene point is actually is actually to the back of the camera which cannot happen right I mean you cannot you cannot image a point which is to the back of the camera so what this means is that if you have a camera facing that way right this is the image plane this is the camera center and if you have a point seen from here then is this point in the front of the camera or to the back of the camera front of the camera right but then I can have a point like this okay like this something like okay now let me draw by actually a different color right and right this goes and this goes and hits it here right now you know here also I get an image point but then right but then it is to the back of the camera right that is what you mean by Z being actually a negative value that means the scene point is coming behind the camera and that does not make sense right so what happens out of this four solutions there is only one solution for which you will get that for both the cameras right not just one because you have a pair for both cameras the point will come in the front it should be positive Z should be positive for both the cameras otherwise you can always get an option where for one camera it turns out to be positive but for the other camera it is negative or it is negative for both the cameras right so basically three of those options get ruled out and using the chirality principle you can you can you can get us you can get us right pin it down to one unique solution that unique solution for which Z is positive right so to arrive at a unique solution to arrive at a unique solution unique solution for of course PR right for PR compute compute the you know 3D points using by say triangulation which I will show next by triangulation by triangulation for each of the four choices each of the four choices of PR and select the one and select the one that reconstructs that reconstructs the point in front of both the cameras in the front of both the cameras. So right this is the reason okay right geometry is more is more interesting right in the sense that it is not so obvious as to what you can do and what you cannot do but then once you start thinking about it makes a lot of sense okay now the point is right now you may you may ask I mean well I may I may not have just one correspondence I may have many right I may supposed to look for look for that configuration right wherein what if what if for something it does not satisfy so what happens reality due to noise right it does not mean that all the points will automatically come in the front for whatever you have the correspondence it does not mean all of them will come in the front so you do what you do is you will take the maximum whichever configuration gives you the maximum number of points in the front you take that okay there is no guarantee that all of them will come in the front because of noise you can end up with some you can end up with no single configuration giving you all the points in the front okay so you take the one that gives you the maximum number of points given the number of correspondences that you have okay and then here the correspondences can be very sparse right because right I mean you just want to you just want to because that is why here you are not asking for a dense reconstruction right this is a sparse correspondence all that I need is I want to find out which one to use okay and to and to and to arrive at that if I can get a get a minimum number of point correspondences which can tell me which R and R and kind of rate T to pick that will give me give me give me the say right sort of the camera configuration.

Now once I have the post is that means that I have I have my R and T right once I have the post then I can come up with another method that can do a dense reconstruction whole idea in structure from motion is get the post first right and to get the post use just minimum number of correspondences that you need okay now the now the next thing that I want to I want to talk about is actually triangulation.